

Measuring consciousness: relating behavioural and neurophysiological approaches

Anil K. Seth¹, Zoltán Dienes², Axel Cleeremans³, Morten Overgaard⁴ and Luiz Pessoa⁵

¹ Department of Informatics, University of Sussex, Falmer, Brighton, BN1 9QJ, UK

² Department of Psychology, University of Sussex, Falmer, Brighton, BN1 9QJ, UK

³ Cognitive Science Research Unit, Université Libre de Bruxelles CP 191, Av. F.-D. Roosevelt 50, 1050 Bruxelles, Belgium

⁴ Cognitive Neuroscience Research Unit, Hammel Neurorehabilitation and Research Center, Aarhus University Hospital, Voldbyvej 15, 8450 Hammel, Denmark

⁵ Department of Psychological and Brain Sciences, Indiana University, 1101 E10th Street, Bloomington, IN 47405, USA

The resurgent science of consciousness has been accompanied by a recent emphasis on the problem of measurement. Having dependable measures of consciousness is essential both for mapping experimental evidence to theory and for designing perspicuous experiments. Here, we review a series of behavioural and brain-based measures, assessing their ability to track graded consciousness and clarifying how they relate to each other by showing what theories are presupposed by each. We identify possible and actual conflicts among measures that can stimulate new experiments, and we conclude that measures must prove themselves by iteratively building knowledge in the context of theoretical frameworks. Advances in measuring consciousness have implications for basic cognitive neuroscience, for comparative studies of consciousness and for clinical applications.

The problem of measurement

How can we measure whether and to what extent a particular sensory, motor or cognitive event is consciously experienced? Such measurements provide the essential data on which the current and future science of consciousness depends, yet there is little consensus on how they should be made. The problem of measuring consciousness differs from the problem of identifying unconscious processing. For instance, in subliminal perception experiments it is desirable to know whether or not a stimulus has been consciously perceived, and in implicit learning paradigms it is interesting to know whether the relationships between different consciously represented stimuli are unconsciously inferred. Measuring consciousness, however, requires saying something about conscious level (Glossary) and conscious content beyond the zero-point of unconsciousness.

Here, we review current approaches for measuring consciousness, covering both behavioural measures and measures based on neurophysiological data. We outline a variety of broad theoretical positions before describing a range of measures in the context of these theories, empha-

sizing recent contributions. We find that potential and actual conflicts among measures suggest new experiments (Table 1); we also assess how different measures can track the graded nature of conscious experience (Table 2). We conclude that it is only by behaving sensibly in a theoretical context that proposed measures pick themselves up by their bootstraps, both validating themselves as measures of what they say they measure and the theories involved.

Theories of consciousness

Worldly discrimination theory

Perhaps the simplest theory that still impacts the experimental literature is that any mental state that can express its content in behaviour is conscious; thus, a person shows they are consciously aware of a feature in the world when they can discriminate it with choice behaviour [1,2]. This theory often makes use of signal-detection theory (SDT), a statistical framework for quantifying the discriminability of a stimulus [3]. SDT itself is mute on the subject of consciousness and can, thus, be combined with

Glossary

Blindsight: the capability of some individuals with visual cortical damage to perform visually guided behaviours even though they report the absence of any associated conscious content [4].

Conscious content: the continually changing phenomenal content (e.g. 'qualia' such as redness and warmth) and intentional content (e.g. explicitly held beliefs, conscious knowledge) that is present to varying degrees at non-zero conscious levels.

Conscious level: applies to a whole organism and refers to a scale ranging from total unconsciousness (e.g. death and coma) to vivid wakefulness. A 'conscious organism' is one that is capable of having non-zero conscious levels. An organism that is dreaming has some conscious level, although the conscious level is reduced in dreamless sleep. A non-zero conscious level indicates the presence of some conscious content.

Primary consciousness: conscious content consisting of a multimodal scene composed of basic perceptual and motor events. Primary consciousness is sometimes called 'sensory consciousness'. By contrast, higher-order consciousness refers to awareness of being in a mental state. In humans it is usually associated with language and an explicit sense of selfhood. In higher-order theories it is possible to have higher-order thoughts that are not themselves (higher-order) conscious, but in virtue of which other (primary) contents are conscious.

Steady-state visual-evoked potential: stimulus-induced components of brain signals that can be identified over extended periods of time. For example a visual image flickering at 10 Hz will evoke a response at 10 Hz in the EEG or magnetoencephalography signal, readily identifiable by a Fourier transform.

Table 1. Conflicts between measures^a

		Content unconscious according to:			
		Objective	Strategic control	Subjective	Wagering
Content conscious according to:	Objective		Unconscious knowledge by Jacoby's process dissociation procedure is <i>ipso facto</i> conscious by objective measures (e.g. Refs [60,72])	In both subliminal perception and implicit learning, subjects often pass objective tasks while claiming to have no knowledge or showing no relation between confidence and accuracy (e.g. [22,73–75])	Shown in blindsight and in the Iowa gambling task by [34]
	Strategic control	Not possible		Subjects can control which grammar to employ while claiming to be guessing [32] and hypnotized subjects can engage in strategic control while reporting no awareness [76]	Not yet shown but entirely possible (see below)
	Subjective	Not possible	Shown in Stroop effects – a person can report the word's meaning but cannot control its rapid use		As yet only shown in our unpublished work – a person can report awareness but still wager indiscriminately
	Wagering	Not possible	Not yet shown but entirely possible (see Box 2)	Not yet shown but entirely possible (see Box 2)	
	Widespread activation	Cognitive control system, including prefrontal cortex, activated by objectively invisible stimuli [56]	Likely for Stroop with clearly shown words [44]	Shown in a 'relative blindsight' paradigm [31]	Likely given the results with verbal subjective measures, but not yet tested
	Synchrony	γ synchrony persists during non-REM sleep and under anaesthesia [53]	As left	As left, also, similar levels of γ synchrony are observed during non-REM and during (reportable) REM sleep [52]	As left
	Complexity measures	Possible in theory (but see Φ). Not tested in practice	As left	As left	As left
Content unconscious according to:					
		Widespread activation	Synchrony	Complexity measures	
Content conscious according to:	Objective	Local neuronal activity can support discriminatory behaviour in many non-conscious organisms (e.g. nematodes and worms). In humans, at least sensory and motor cortices need to be active	Unlikely given current evidence	Not yet tested but entirely possible	
	Strategic control	Unlikely: strategic control probably requires activation in both perceptual and frontal regions	As above	Possible but not tested	
	Subjective	As above	As above	As above	
	Wagering	As above	As above	As above	
	Widespread activation		Experimentally open. Some studies show increased long-range synchrony accompanying conscious access [48]	Possible in theory. Not tested in practice	
Synchrony	Gamma synchrony is often localized [47]		As above		
Complexity measures	High neural complexity (or Φ , or c_d) probably requires widespread activity: all else being equal, larger networks will give rise to higher complexity values [17]	Possible in theory. Not tested in practice			

^aRows indicate a measure finds the content conscious and columns indicate the measure finds the content unconscious. Entries reflect a scale according to which a particular conflict is (i) experimentally noted, (ii) not yet shown but entirely possible, (iii) experimentally open, (iv) possible but not tested, (v) unlikely given current evidence and/or theory, or (vi) not possible.

different theories. The combination of SDT with the worldly discrimination theory (WDT) asserts that continuous information available for discriminations is necessarily the content of conscious mental states. This theory captures one property of conscious knowledge, namely that it enables choice behaviour. However, rightly or wrongly, it does not respect other properties that are associated with consciousness. For example, according to this theory blindsight patients see consciously because forced-choice discrimination is the result by which we infer

that they can see at all. However, two properties of blindsight indicate intuitively that the seeing is not conscious [4]. First, blindsight patients do not spontaneously attempt to use the information practically or inferentially. Second, blindsight patients themselves think they cannot see.

Integration theories

Other theories attempt to locate a divide between conscious and unconscious processes that respect one or both

Table 2. Sensitivity to graded consciousness^a

Type	Measure	Primary theoretical affiliation	Sensitivity to graded conscious level	Sensitivity to graded conscious content
Behavioural				
Objective	Discrimination behaviour	WDT	None (either an organism is sufficiently conscious to show choice behaviour, or it is not)	The d' value in SDT might index graded consciousness, though typically any $d' > \text{zero}$ is taken to imply full consciousness [1]
Objective	Strategic control	Integration theory	None (see above)	None so far. Various equations developed assume that a content is either clearly conscious or unconscious (e.g. Ref. [9])
Subjective	Introspective report	HOT	Poor and indirect; poor verbal coherence might indicate low conscious level	Introspective reports are explicitly highly sensitive to conscious content and can indicate close mismatches between observed and reported states
Subjective	Confidence ratings	HOT	Poor and indirect; confidence might diminish with conscious level	Confidence can indicate degrees of higher-order belief
Subjective	PDW	HOT	Poor and indirect though various continuous measures can be used [35]	Gambling measures can indicate degrees of higher order belief (see Box 2)
Objective and subjective	Glasgow coma scale	None	High	None
Neurophysiological				
EEG	Bispectral index	None	High	None
EEG/MEG	Early ERP ('awareness negativity' [77])	Localized integration [14,39]	Most ERPs are attenuated by sleep and low arousal, but yet not directly tested for awareness negativity	Some. Early ERPs are delayed for low-contrast stimuli [77]
EEG/MEG	Late ERP (P300)	Global Integration [40]	P300 can be elicited during sleep though with different profile [78]	Low. P300 dichotomously characterizes 'seen' versus 'not seen' trials [40]
General neuroimaging	Widespread activation	Integration	Imaging of consciousness-impaired patients can distinguish different conscious levels [45]	Low. Access to global workspace is usually considered all-or-none [10]
Synchrony	Induced γ activity	Integration (local and/or global)	Synchrony is present even in non-REM sleep [53]	Not tested (to our knowledge)
Synchrony	SSVEP (frequency 'tag')	Global integration	Auditory frequency tag is modulated by arousal level [79]	Not tested (to our knowledge)
Complexity	Neural complexity	Integration	High (in principle but not yet shown)	Low
Complexity	Information integration (Φ)	Integration	High (in principle but not yet shown)	Some (in principle Φ can gauge conscious contents)
Complexity	Causal density	Integration	High (in principle; shown only in our own unpublished work)	Possibly revealed by causal interaction patterns but not yet shown

^aConscious level can be graded on a scale from coma to full wakefulness, and conscious contents can also be graded (e.g. fringe consciousness and Ganzfeld experiences). This table indicates how different measures are able to track graded consciousness, as well as their primary theoretical affiliation.

of the intuitions just mentioned. According to integration theories, conscious contents are widely available to many cognitive and/or neural processes. This core idea has been expressed in various ways. In philosophy, it has been described as inferential promiscuity [5], fame in the brain [6], the unified field theory [7] and global access; in cognitive psychology as broadcast within a global workspace [8] and in a more constrained way as the process dissociation framework [9]; and in neuroscience as a neuronal global workspace [10], a dynamic core [11,12], integrated information [13], and, in a more constrained way, as locally recurrent activity [14] or neuronal synchrony [15,16]. The neuroscience theories in particular have given rise to several putative measures that have been used to quantify simultaneous integration and differentiation in neural dynamics on the basis that conscious experience is also simultaneously integrated and differentiated [17]. According to these theories a mental state is conscious if it provides a sufficiently informative discrimination among a large repertoire of possible states, in which successful discrimination requires both differentiation and integration [11,12].

Higher-order thought theories

According to higher-order thought (HOT) theories, a mental state is conscious when a person is actually aware [18] or disposed toward being aware [19] of being in that state. Theories differ according to whether awareness of the mental state is achieved by perceiving it [20] or thinking about it [18]. HOT theories differ from WDTs in that it is the ability of a person to discern their mental state, rather than the state the world is in, which determines whether a mental state is conscious. In the context of the SDT, HOT theories is associated either with the criterion of standard SDT or with the second level of discrimination – discriminating not the world (as in the WDT) but the accuracy of our responses [21].

Because of their differing theoretical affiliations, measures of consciousness can, and do, conflict with each other, as detailed in Table 1. Also, measures of consciousness not only should distinguish between conscious and unconscious processing but also indicate the degree to which an organism or a mental state is conscious [22,23]. Sensitivity to graded consciousness is described in Table 2. All theories described so far, with the exception

of some neural integration theories [11,13], describe conditions for asserting whether a particular mental state is conscious (conscious content). They do not generally pertain to whether an organism is conscious or unconscious at a particular time (conscious level). As we will see, measures of consciousness can, and do, address both of these issues.

Behavioural measures

'Objective measures' assume the ability to choose accurately under forced choice conditions as indicating a conscious mental state [24,25]. For example being able to pick which item might come next indicates conscious knowledge of regularities in sequences. Conversely, knowledge is unconscious if a distinction in the world expresses itself only in non-intentional characteristics of behaviour (such as its speed) or in galvanic skin response, functional magnetic resonance imaging (fMRI) or other physiological characteristics not expressed in behaviour at all [26]. That is, knowledge is unconscious if it expresses itself in an indirect – but not a direct – test [27,28]. Unqualified trust in objective measures presupposes WDT and conflicts with most other measures (Table 1).

'Strategic control' determines the conscious status of knowledge by the person's ability to deliberately use or not use the knowledge according to instructions. In Jacoby's process dissociation procedure [9], a person either tries to avoid using the information (exclusion task) or makes sure they do use it (inclusion task); any difference in influence of the stimulus between these conditions indicates conscious knowledge, and any use of it despite intentions in the exclusion condition indicates unconscious knowledge (e.g. Refs [29,30]). Unqualified trust in this measure presupposes integration theory.

'Subjective measures' require subjects to report their mental states. Most simply, subjective measures have been used to ascertain whether a person knows that they know. The WDT indicates knowledge but not the awareness of knowing. To test for awareness of knowing, confidence ratings can be given. If for all the trials when the person says 'guess', the discrimination performance is still above baseline, then there is evidence that the person has knowledge that they do not know they have: unconscious knowledge by the 'guessing criterion' [31]. If a person's knowledge states are conscious, they will know when they know and when they are just guessing. In this case, there should be a relationship between confidence and accuracy, indicating conscious knowledge with no relationship indicating unconscious knowledge by the 'zero-correlation criterion' [32,33]. Unqualified trust in subjective measures presupposes one or other of the higher-order theories.

An advantage of subjective measures is that the conscious status of a range of mental states can be assessed, including both knowledge content and phenomenal content (Box 1). For example a blindsight patient can consciously know without consciously seeing – if they think they know but they do not think they see. Graded degrees of conscious seeing were assessed by Overgaard *et al.* [22]: normal subjects consistently reported glimpses or impressions of content they were not willing to say they actually saw (Table 2). Subjective specification of conscious content is often associated with introspection, but not all subjective reports are

Box 1. Structural knowledge and judgment knowledge

Tasks can involve a range of knowledge states, the conscious status of each can be assessed subjectively. For example, when a person is exposed to a structured domain (e.g. strings from an artificial grammar), they learn about the structure (structural knowledge). Artificial grammar learning might consist of the knowledge that an 'M' can start a string, about whole strings that were presented and so on. In the test phase, the structural knowledge is brought to bear on a test item to form a new piece of knowledge: the judgment, for example, that this string is grammatical (judgment knowledge) [58]. Structural knowledge can be unconscious when judgment knowledge is conscious. For example in natural language you can consciously know whether a sentence in your native tongue is grammatical or not (conscious judgment knowledge) but have no idea why you know that. It is important to be clear whether a measure tests the conscious status of judgment or structural knowledge. Confidence ratings and wagering measures involve confidence or wagers on a judgment, therefore the guessing and zero-correlation criteria in these cases test the conscious status of judgment knowledge only. Similarly, Jacoby's process dissociation procedure measures the conscious status of judgment knowledge [9]: in implicit learning tasks for example, a person can exclude effectively because they consciously know that a response satisfies structural constraints without consciously knowing what those structural constraints are (e.g. that the response forms part of a long-distance dependency, of a symmetry and so on) [59,60].

Dienes and Scott [58] introduced a simple subjective way of measuring the conscious status of structural knowledge: for each judgment in a test phase, subjects indicated whether their judgment was based on random guessing, intuition, conscious rules or memory. Guessing and intuition *prima facie* indicate unconscious structural knowledge, and conscious rules and memory indicate conscious structural knowledge. Dienes [21] argued that structural knowledge might be the interesting target for implicit learning research (insofar as it has indicated qualitative differences in knowledge), and that judgment knowledge might be the interesting target for perception research.

introspective given that introspection requires being consciously aware of being in a mental state (rather than merely being consciously aware of states in the world) [18,22].

Post-decision wagering

Post-decision wagering (PDW) is a recent variation on confidence ratings whereby subjects make a first-order discrimination and then place a wager (rather than a confidence rating) regarding the outcome of the discrimination [34,35]. As with confidence ratings PDW presupposes a version of the HOT theory. Yet, because PDW does not ask for subjective reports, its proponents claim that it is a direct and objective measure of consciousness (see Box 2 for arguments against this claim). An advantage is that the lack of subjective reports enables the method to be used with children [35] and animals.

Finally, most behavioural measures are aimed at assessing whether particular mental content is conscious, not whether an organism is conscious. One exception is the Glasgow coma scale, a set of behavioural tests used to assess the presence, absence and depth of coma in patients with brain trauma [36]. In clinical contexts such behavioural tests are increasingly being augmented by brain-based measures of conscious level.

Brain Measures

Electroencephalegraphic measures

In 1929, Hans Berger discovered that waking consciousness is associated with low-amplitude, irregular electroencepha-

Box 2. Post-decision wagering: a 'direct' measure of awareness?

In PDW, subjects make a first-order discrimination and then place a wager on its outcome [34]. Unconscious knowledge can be shown by above-chance first-order discriminations when (i) low wagers are given (guessing criterion) or (ii) there is no relationship between wagering and accuracy (zero correlation criterion). In one example the blindsight subject 'GY' classified a sensory stimulus as either present or absent, and then wagered either a small monetary stake or a large stake on the correctness of this classification. Although GY made the correct classification on ~70% of trials, he was just as likely to bet low as he was to bet high on these trials. This absence of advantageous wagering is taken as evidence for absence of consciousness of the correctness of the first-order discrimination. Conversely, good first-order performance accompanied by advantageous wagering is taken as evidence of awareness of the first-order stimuli. Like confidence ratings, PDW requires the subject to make a 'metacognitive' judgment about a (putatively) conscious experience but it differs by implementing this requirement indirectly, via a wager. As a result there can be conflicts between wagering and verbal reports that directly express HOTs (see Table 1 in the main text). Because wagering might avoid some biases affecting introspective and confidence reports (e.g. subjects can be reluctant to report weakly perceived stimuli [22]), PDW has been asserted to provide a 'direct and objective' measure of awareness [34]. This is a strong claim that is difficult to justify [61–63]:

- All behavioural measures, including PDW, require a response criterion: for example, whether to push a button or not (therefore claiming a 'direct' behavioural measure might be mistaken from the outset). Any response criterion can be affected by cognitive bias, and, for PDW, a plausible bias could arise from risk aversion. As with confidence methods, the zero correlation criterion can take account of bias but any trial-to-trial variation in bias will still undermine its sensitivity.
- Because PDW does not ask for subjective reports, it is difficult to exclude the possibility that advantageous wagering could be learned unconsciously. This could be shown by wagering advantageously (based on unconscious judgment knowledge) while always believing that one is guessing.
- Because PDW requires a metacognitive judgment about a putatively conscious experience, it is apparently no more 'objective' than a confidence judgment.

PDW highlights the interdependence of measures and theories. According to HOT theories the metacognitive nature of PDW is not problematic because some metacognitive content is constitutive of any conscious state. However, from non-HOT perspectives, the absence of wagering-related metacognitive content does not by itself imply the absence of primary (sensory) conscious content.

lographic (EEG) activity in the 20–70 Hz range. It is now known that unconscious conditions such as non-REM sleep, coma, general anaesthesia and epileptic absence seizures show predominantly low-frequency, regular and high-amplitude oscillations [37,38]. Event-related cortical potentials (ERPs) have been used to assess whether a stimulus is consciously perceived or not, although there is dispute about whether early (e.g. 'visual awareness negativity', ~100 ms [39]) or late (e.g. the 'P300', ~300 ms [40]) components are most relevant. ERPs also are associated with other functions beyond consciousness *per se*, for example in novelty detection or recognition [41]. The proprietary 'bispectral index' (BIS) combines various aspects of the EEG signal to estimate anaesthetic depth (conscious level) and hence the probability of accidental waking during surgery [42]. EEG measures either float free of theory, gaining purchase through reliable correlation (e.g. BIS), or assume a version of integration theory in which the appearance of a particular

ERP indicates global availability [40] or locally recurrent processing [39] (Table 2).

Widespread activation

In line with integration theories, abundant evidence indicates that consciously perceived inputs elicit widespread brain activation, as compared with inputs that do not reach consciousness [43]. For example, Dehaene and colleagues have shown in a visual masking paradigm that consciously seen stimuli activate a broad frontoparietal network compared with unseen stimuli, by using both fMRI [44] and ERP signals [40]. Neuroimaging of vegetative and minimally conscious patients also reveals stimulus-evoked activity only in sensory cortices [45]. However, differences in conscious perception are often confounded with differences in performance. Lau and Passingham [31] controlled for this confound by using a metacontrast masking paradigm and found that conscious and unconscious conditions are differentiated only by activity in the left mid-dorsolateral prefrontal cortex; widespread brain activity was found in both conditions given sufficiently accurate performance. These results indicate that widespread activation can conflict with other measures (Table 1), although it is difficult to know whether the additional prefrontal activity is related to the generation of conscious content and/or to subjective report of that content.

Synchrony

Several researchers have suggested that consciousness arises from transient neuronal synchrony, possibly in the γ (30–70 Hz) [15,16] or β (~15 Hz) [46] ranges. Measuring consciousness by synchrony presupposes integration theories of at least a limited kind (to the extent that local synchrony is deemed sufficient [14,47]). Several studies have reported an association between synchrony and consciousness, both in induced γ -range activity [47,48] and in steady-state visual-evoked potentials (SSVEP) ('frequency tags' [49]). However, there is not yet evidence that disruption of γ -band synchrony leads to disruption of conscious contents [50], and γ oscillations have been associated with a wide range of cognitive functions in addition to consciousness *per se* [51]. Moreover, γ synchrony can be present equally during REM (consciously vivid) and non-REM (dreamless) sleep states [52], and also can be high during anaesthesia [53]. Together these observations indicate that neuronal synchrony might at best be necessary but that is not sufficient for consciousness.

Complexity, information integration and causal density

Several recent measures build on the observation that conscious scenes are distinguished by being simultaneously integrated (each conscious scene is experienced 'all of a piece') and differentiated (each conscious scene is composed of many distinguishable components and is therefore different from every other conscious scene) [11,13,17] (Box 3). The dynamic core hypothesis (DCH) proposes that consciousness arises from neural dynamics in the thalamocortical system with just these features, as measured by the quantity 'neural complexity' (C_N). C_N is an information-theoretic measure; the C_N value is high if each subset of a neural system can take on many different states

Box 3. Consciousness and complexity

Three recently proposed measures – neural complexity C_N [54], information integration Φ [13] and causal density c_d [55] – attempt to capture the coexistence of integration and differentiation that is central to ‘complexity’ theories of consciousness (Figure 1). All these measures are defined in terms of the stationary dynamics of a neural system (X), composed of N elements. C_N and Φ are based on information theory, whereas c_d is based on multivariate autoregressive modelling. The neural complexity $C_N(X)$ of X is calculated as the average mutual information (MI); a measure of statistical dependence) among subsets of all possible sizes for all bipartitions of X . This quantity is high if small subsets of X show high statistical independence but large subsets show low independence. In view of the computational expense of considering all bipartitions, C_N can be approximated by considering only bipartitions of one element and the remainder of the system; another approximation derives directly from network topology [64].

Information integration $\Phi(X)$ is defined as the effective information (EI) across the ‘minimum information bipartition’ (MIB) of X , where EI is a directed version of MI that depends on stimulating one half of a bipartition with random (maximally entropic) activity and where the MIB is that bipartition for which the normalized EI is minimum, the informational ‘weakest link’ [65]. Whereas C_N is a measure of actual neural activity, Φ is a measure of the capacity of a system to integrate information. Like C_N , Φ is infeasible to compute for large N . It is also obviously challenging to inject arbitrary subsets of real biological systems with random activity.

Causal density $c_d(X)$ is calculated as the fraction of interactions among X ’s elements that are causally significant, as identified by ‘Granger causality’ [66]. Granger causality is a statistical interpretation of causality in which x_1 ‘causes’ x_2 if knowing the past of x_1 helps predicts the future of x_2 better than knowing the past of x_2 alone. It is usually calculated by linear autoregression, although nonlinear extensions exist [67]. High c_d indicates that elements in X are both globally coordinated (to be useful for predicting the activities of others) and dynamically distinct (to contribute to these predictions in different ways). Like Φ but not C_N , c_d is sensitive to causal interactions in neural dynamics. Like C_N but not Φ , it reflects the activity and not the capacity of X . Like both, it is difficult to calculate for large N .

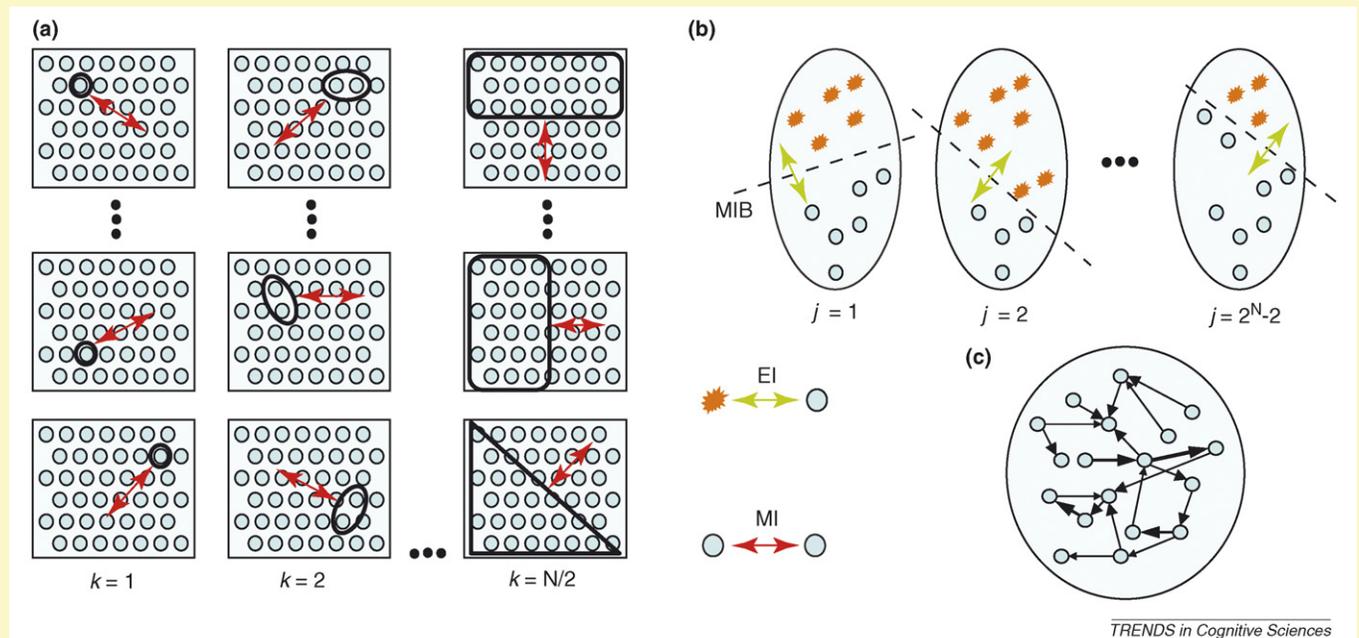


Figure 1. Measuring integration and differentiation in neural dynamics, for a system composed of N elements. (a) C_N is calculated as the sum of the average MI over $N/2$ sets of bipartitions indexed by k (e.g. for $k=1$ an average MI is calculated over N bipartitions). (b) Φ is calculated as the EI across the MIB. To calculate EI for a given bipartition (indexed by j), one subset is injected with maximally entropic activity (orange stars) and MI across the partition is measured. (c) c_d is calculated as the fraction of interactions that are causally significant according to Granger causality. A weighted (and unbounded) version of c_d can be calculated as the summed magnitudes of all significant causal interactions (depicted by arrow width).

TRENDS in Cognitive Sciences

and if these states make a difference to the rest of the system [54].

The information integration theory of consciousness (IITC) shares with the DCH the idea that conscious experiences provide informative discriminations among a vast repertoire of possible experiences [13]. In the IITC, the quantity Φ is defined as the information that is integrated across the informational ‘weakest link’ of a system. Importantly, Φ is a measure of the capacity of a neural system to integrate information, whereas C_N is a measure of the actual dynamics of the system. A third measure, causal density (c_d), measures the fraction of causal interactions among elements of a system that are statistically significant; it is low both for highly integrated systems and for collections of independent elements [55].

Unqualified trust in C_N , Φ or c_d presupposes an integration theory. This is particularly explicit for Φ because the

IITC defines consciousness as information integration, implying that high Φ in any system, biological or otherwise, is sufficient for consciousness. Although all three measures are well grounded in theory, in practice they are difficult to measure, and their experimental exploration stands as an important challenge.

Measures, theories and conflicts

Theories of consciousness recommend the use of certain measures, and the use of certain measures presuppose particular theories. Just as theoretical positions conflict with one another, conflicts among measures can be expected and, in many cases, have been observed (Table 1). These conflicts can guide further experiments and theoretical refinements. For example the extent to which PDW corresponds with other behavioural measures will shed light on whether wagering involves separate

Review

Box 4. Outstanding questions

- Can the neural mechanisms underlying subjective report be dissociated from those underlying consciousness *per se* [14,68]?
- Which possible conflicts between measures indicated in Table 1 (in the main text) can be demonstrated? Which measures cohere together? Under what conditions do the answers produced by a measure make theoretical sense?
- How can multiple measures be combined to better isolate the neural mechanisms of consciousness? Can multiple measures isolate independent processes underlying conscious experience?
- Do C_N , Φ or c_d behave as predicted by theory? Answering this question depends on (i) experimental methods of sufficient spatiotemporal resolution to reveal relevant details of thalamo-cortical activity, and (ii) sensible approximations enabling application to large neural datasets.
- Can a theoretically principled objective measure improve on current clinical methods of diagnosing anaesthesia and impaired consciousness after brain injury?
- How does a measure of consciousness affect what it supposedly measures? This question relates to behavioural subjective methods, especially introspection [69].
- Which measures can be applied to infants and non-human animals and how should the results be interpreted [70,71]?

mechanisms of higher-order access, potentially indicating new aspects of HOT theories. Regarding brain measures, results indicating the insufficiency of widespread activation [31,56] and γ synchrony [52,53] (when conscious contents are measured by subjective report) challenge basic integration theories and indicate that new insights will be uncovered by comparing these measures with those based on complexity theory.

The most informative new studies will be those that combine multiple measures, both behavioural and brain-based (Box 4). Presently, these measures tend to pick up on different aspects of consciousness: behavioural measures are mostly used for assessing which contents are conscious, whereas at least some brain-based measures seem well suited for measuring conscious level; graded consciousness can, in principle, be assessed by either type but in different ways (Table 2). Therefore, an integrative approach combining both types of measure in a single study encourages a virtuous circularity in which putative measures and theoretical advances mutually inform, validate and refine one another. The ultimate virtue in a measure is not its *a priori* toughness, but its ability to build on intuitions, identify interesting divides in nature and then correct the foundations on which it was built [57].

Acknowledgements

A.C. is supported by Concerted Research Action 06/11-342 (Belgium). A.C. and M.O. are supported by European Commission FP6 Grant #043457 'Mindbridge: Measuring Consciousness'. L.P. is supported in part by the National Institute of Mental Health, USA (1R01 MH071589).

References

- 1 Dulany, D.E. (1997) Consciousness in the explicit (deliberative) and implicit (evocative). In *Scientific Approaches to Consciousness* (Cohen, J. and Schooler, J., eds), pp. 179–211, Lawrence Erlbaum Associates
- 2 Eriksen, C.W. (1960) Discrimination and learning without awareness: a methodological survey and evaluation. *Psychol. Rev.* 67, 279–300
- 3 Green, D.M. and Swets, J.A. (1966) *Signal Detection Theory*, Wiley
- 4 Weiskrantz, L. (1998) *Blindsight: A Case Study and Implications*, Oxford University Press
- 5 Stich, S. (1978) Beliefs and subdoxastic states. *Philos. Sci.* 45, 499–518
- 6 Dennett, D. (2005) *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*, MIT Press
- 7 Searle, J. (2004) *Mind: A Brief Introduction*, Oxford University Press
- 8 Baars, B.J. (1988) *A Cognitive Theory of Consciousness*, Cambridge University Press
- 9 Jacoby, L. (1991) A process dissociation framework: separating automatic from intentional uses of memory. *J. Mem. Lang.* 30, 513–541
- 10 Dehaene, S. et al. (2003) A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proc. Natl. Acad. Sci. U. S. A.* 100, 8520–8525
- 11 Tononi, G. and Edelman, G.M. (1998) Consciousness and complexity. *Science* 282, 1846–1851
- 12 Edelman, G.M. (2003) Naturalizing consciousness: a theoretical framework. *Proc. Natl. Acad. Sci. U. S. A.* 100, 5520–5524
- 13 Tononi, G. (2004) An information integration theory of consciousness. *BMC Neurosci.* 5, 42
- 14 Lamme, V.A. (2006) Towards a true neural stance on consciousness. *Trends Cogn. Sci.* 10, 494–501
- 15 Engel, A.K. et al. (1999) Temporal binding, binocular rivalry, and consciousness. *Conscious. Cogn.* 8, 128–151
- 16 Crick, F. and Koch, C. (1990) Towards a neurobiological theory of consciousness. *Semin. Neurosci.* 2, 263–275
- 17 Seth, A.K. et al. (2006) Theories and measures of consciousness: an extended framework. *Proc. Natl. Acad. Sci. U. S. A.* 103, 10799–10804
- 18 Rosenthal, D.M. (2005) *Consciousness and Mind*, Clarendon
- 19 Carruthers, P. (1996) *Language, Thought, and Consciousness*, Cambridge University Press
- 20 Lycan, W.G. (2004) The superiority of HOP over HOT. In *Higher-Order Theories of Consciousness: An Anthology* (Gennaro, R.J., ed.), pp. 93–113, John-Benjamins
- 21 Dienes, Z. (2008) Subjective measures of unconscious knowledge. In *Models of Brain and Mind: Physical, Computational and Psychological Approaches* (Banerjee, R. and Chakrabarti, C., eds), pp. 49–64, Elsevier
- 22 Overgaard, M. et al. (2006) Is conscious perception gradual or dichotomous? A comparison of report methodologies during a visual task. *Conscious. Cogn.* 15, 700–708
- 23 Cleeremans, A. (2006) Conscious and unconscious cognition: a graded, dynamic, perspective. In *Progress in Psychological Science Around the World* (Vol. 1) (Jing, Q. et al., eds), In pp. 401–418, Psychology Press
- 24 Pessoa, L. et al. (2006) Target visibility and visual awareness modulate amygdala responses to fearful faces. *Cereb. Cortex* 16, 366–375
- 25 Smyth, A. and Shanks, D.R. (2008) Awareness in contextual cueing with extended and concurrent explicit tests. *Mem. Cognit.* 36, 403–415
- 26 Naccache, L. et al. (2005) A direct intracranial record of emotions evoked by subliminal words. *Proc. Natl. Acad. Sci. U. S. A.* 102, 7713–7717
- 27 Reingold, E.M. and Merikle, P.M. (1990) On the inter-relatedness of theory and measurement in the study of unconscious processes. *Mind Lang.* 5, 9–28
- 28 Greenwald, A.G. et al. (2003) Long-term semantic memory versus contextual memory in unconscious number processing. *J. Exp. Psychol. Learn. Mem. Cogn.* 29, 235–247
- 29 Norman, E. et al. (2007) Gradations of awareness in a modified sequence learning task. *Conscious. Cogn.* 16, 809–837
- 30 Destrebecqz, A. and Cleeremans, A. (2003) Temporal effects in sequence learning. In *Attention and Implicit Learning* (Jimenez, J.C., ed.), pp. 181–213, John Benjamins
- 31 Lau, H.C. and Passingham, R.E. (2006) Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proc. Natl. Acad. Sci. U. S. A.* 103, 18763–18768
- 32 Dienes, Z. et al. (1995) Unconscious knowledge of artificial grammars is applied strategically. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 1322–1338
- 33 Kolb, F.C. and Braun, J. (1995) Blindsight in normal observers. *Nature* 377, 336–338
- 34 Persaud, N. et al. (2007) Post-decision wagering objectively measures awareness. *Nat. Neurosci.* 10, 257–261
- 35 Ruffman, T. et al. (2001) Does eye gaze indicate knowledge of false belief? *J. Exp. Child Psychol.* 80, 201–224
- 36 Teasdale, G.M. and Murray, L. (2000) Revisiting the Glasgow coma scale and coma score. *Intensive Care Med.* 26, 153–154
- 37 Berger, H. (1929) Ueber das Elektroenkephalogramm des Menschen. *Archiv fuer Psychiatrie und Nervenkrankheiten, Berlin* 87, 527–570

- 38 Baars, B.J. *et al.* (2003) Brain, conscious experience and the observing self. *Trends Neurosci.* 26, 671–675
- 39 Koivisto, M. *et al.* (2006) Independence of visual awareness from the scope of attention: an electrophysiological study. *Cereb. Cortex* 16, 415–424
- 40 Del Cul, A. *et al.* (2007) Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biol.* 5, e260
- 41 Donchin, E. and Coles, M. (1988) Is the P300 component a manifestation of context updating? *Behav. Brain Sci.* 11, 357–374
- 42 Myles, P.S. *et al.* (2004) Bispectral index monitoring to prevent awareness during anaesthesia: the B-aware randomised controlled trial. *Lancet* 363, 1757–1763
- 43 Baars, B.J. (2002) The conscious access hypothesis: origins and recent evidence. *Trends Cogn. Sci.* 6, 47–52
- 44 Dehaene, S. *et al.* (2001) Cerebral mechanisms of word masking and unconscious repetition priming. *Nat. Neurosci.* 4, 752–758
- 45 Laureys, S. (2005) The neural correlate of (un)awareness: lessons from the vegetative state. *Trends Cogn. Sci.* 9, 556–559
- 46 Gross, J. *et al.* (2004) Modulation of long-range neural synchrony reflects temporal limitations of visual attention in humans. *Proc. Natl. Acad. Sci. U. S. A.* 101, 13050–13055
- 47 Palva, S. *et al.* (2005) Early neural correlates of conscious somatosensory perception. *J. Neurosci.* 25, 5248–5258
- 48 Melloni, L. *et al.* (2007) Synchronization of neural activity across cortical areas correlates with conscious perception. *J. Neurosci.* 27, 2858–2865
- 49 Srinivasan, R. *et al.* (1999) Increased synchronization of magnetic responses during conscious perception. *J. Neurosci.* 19, 5435–5448
- 50 Rees, G. *et al.* (2002) Neural correlates of consciousness in humans. *Nat. Rev. Neurosci.* 3, 261–270
- 51 Buzsaki, G. (2006) *Rhythms of the Brain*, Oxford University Press
- 52 Bullock, T.H. *et al.* (1995) Temporal fluctuations in coherence of brain waves. *Proc. Natl. Acad. Sci. U. S. A.* 92, 11568–11572
- 53 Vanderwolf, C.H. (2000) Are neocortical gamma waves related to consciousness? *Brain Res.* 855, 217–224
- 54 Tononi, G. *et al.* (1994) A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proc. Natl. Acad. Sci. U. S. A.* 91, 5033–5037
- 55 Seth, A.K. (2005) Causal connectivity analysis of evolved neural networks during behavior. *Network.* 16, 35–54
- 56 Lau, H.C. and Passingham, R.E. (2007) Unconscious activation of the cognitive control system in the human prefrontal cortex. *J. Neurosci.* 27, 5805–5811
- 57 Chang, H. (2004) *Inventing Temperature: Measurement and Scientific Progress*, Oxford University Press
- 58 Dienes, Z. and Scott, R. (2005) Measuring unconscious knowledge: distinguishing structural knowledge and judgment knowledge. *Psychol. Res.* 69, 338–351
- 59 Fu, Q. *et al.* (2008) Implicit sequence learning and conscious awareness. *Conscious Cogn.* 17, 185–202
- 60 Destrebecqz, A. and Cleeremans, A. (2001) Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychon. Bull. Rev.* 8, 343–350
- 61 Seth, A.K. (2007) Post-decision wagering measures metacognitive content, not sensory consciousness. *Conscious Cogn.* DOI: 10.1016/j.concog.2007.05.008 (www.sciencedirect.com)
- 62 Persaud, N. *et al.* (2007) Reply to note by Seth: experiments show what post-decision wagering measures. *Conscious Cogn.* DOI: 10.1016/j.concog.2007.06.002 (www.sciencedirect.com)
- 63 Seth, A.K. (2007) Theories and measures of consciousness develop together. *Conscious. Cogn.* DOI: 10.1016/j.concog.2007.08.004 (www.sciencedirect.com)
- 64 De Lucia, M. *et al.* (2005) A topological approach to neural complexity. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 71, 016114
- 65 Tononi, G. and Sporns, O. (2003) Measuring information integration. *BMC Neurosci.* 4, 31
- 66 Ding, M. *et al.* (2006) Granger causality: basic theory and application to neuroscience. In *Handbook of Time Series Analysis* (Schelter, S. *et al.*, eds), pp. 438–460, Wiley
- 67 Ancona, N. *et al.* (2004) Radial basis function approaches to nonlinear Granger causality of time series. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 70, 056221
- 68 Block, N. (2007) Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behav. Brain Sci.* 30, 481–548
- 69 Overgaard, M. *et al.* (2006) The electrophysiology of introspection. *Conscious Cogn.* 15, 662–672
- 70 Seth, A.K. *et al.* (2005) Criteria for consciousness in humans and other mammals. *Conscious Cogn.* 14, 119–139
- 71 Washburn, D.A. *et al.* (2006) Rhesus monkeys (*Macaca mulatta*) immediately generalize the uncertain response. *J. Exp. Psychol. Anim. Behav. Process* 32, 185–189
- 72 Debner, J.A. and Jacoby, L.L. (1994) Unconscious perception: attention, awareness, and control. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 304–317
- 73 Cheesman, J. and Merikle, P.M. (1984) Priming with and without awareness. *Percept. Psychophys.* 36, 387–395
- 74 Dienes, Z. and Longuet-Higgins, H.C. (2004) Can musical transformations be implicitly learned? *Cogn. Sci.* 28, 531–558
- 75 Szczepanowski, R. and Pessoa, L. (2007) Fear perception: can objective and subjective awareness measures be dissociated? *J. Vis.* (Vol. 4) 7, 10
- 76 Dienes, Z. and Perner, J. (2007) The cold control theory of hypnosis. In *Hypnosis and Conscious States: The Cognitive Neuroscience Perspective* (Jamieson, G., ed.), pp. 293–314, Oxford University Press
- 77 Wilenius, M.E. and Revonsuo, A.T. (2007) Timing of the earliest ERP correlate of visual awareness. *Psychophysiology* 44, 703–710
- 78 Colrain, I.M. and Campbell, K.B. (2007) The use of evoked potentials in sleep research. *Sleep Med. Rev.* 11, 277–293
- 79 Griskova, I. *et al.* (2007) The amplitude and phase precision of 40 Hz auditory steady-state response depend on the level of arousal. *Exp. Brain Res.* 183, 133–138